

Solving the problem of hydrometeorological data analysis by random process

Dương Ton Dam, Dang Kien Cuong

Abstract—The paper reviews the hydrometeorological data analysis (precipitation, flow, water level, etc.) to evaluate and predict mutations such as flood, drought and saline extremes to reduce the impact of climate change on the economy and life. The main method for solving the problem posed is the max-domain of attraction of extreme distributions with the Gumbel copula of random variables related to hydrometeorological data. Results presented in this paper are reviewed and verified through data supplied by hydrometeorological stations at the Tan Chau and Chau Doc districts, An Giang province from 1990 to present.

Index Terms—Copula, Extreme Value distributions, Maximal domain of attraction, Method of maximum likelihood, Newton – Raphson algorithm.

1 INTRODUCTION

Probability and statistic theories have been widely applied in hydrometeorological analysis. In this review, we introduce the integration of random (or stochastic) process methods to evaluate, predict and analyze the extreme distributions in hydrometeorology. In other words, this review will be solved problems associated with flood peak, drought and saline extremes, contributing the significant economy and life.

The mathematical methods are mainly used in

Manuscript Received on July 13th, 2016. Manuscript Revised December 06th, 2016.

This research is funded by Vietnam National University Ho Chi Minh City (VNU-HCM) under grant number C2017-26-03.

Duong Ton Dam was Vice Rector of University of Information Technology, Vietnam National University, Ho Chi Minh City 2007 – 2012. He is currently General Secretary of the Mathematical Society Ho Chi Minh City.

Dang Kien Cuong is currently pursuing the Ph.D. degree in computer science University of Information Technology, Vietnam National University, Ho Chi Minh City. He is currently the lecture at Nong Lam University, Ho Chi Minh City.

this paper as follows:

- ✓ Method of moments and Maximum Likelihood Estimation (MLE).
- ✓ The probability distribution and random variables (expectation, covariance, correlation, ...).
- ✓ The copula method offers a efficiency for describing the dependence among multiple hydrologic variables.

In the last section of the research, the data of maximum rainfall and water levels are treated in Tan Chau district, An Giang province, located in the Mekong River Delta of Vietnam.

2 THEORY BACKGROUND

2.1 Extreme Value distributions

Let X_1, X_2, \dots, X_n be a sequence of independent random variable. When we consider the max value of this sequence, $M_n = \max \{X_1, X_2, \dots, X_n\}$ has cumulative distribution function:

$$F_{M_n}(x) = P(M_n \leq x) = P\{X_1 \leq x, \dots, X_n \leq x\} \\ = P\{X_1 \leq x\} \dots P\{X_n \leq x\} = F_{X_1}(x) \dots F_{X_n}(x)$$

If random sequence, $X_i; i = 1, 2, \dots, n$; independent and they have the same distribution $F_X(x)$, we get:

$$F_{M_n}(x) = (F_X(x))^n$$

The symbol of density function equivalence with $F_X(x)$ is $f_X(x) = (F_X(x))'$. Then we will determine the density function of $M_n = \max \{X_1, X_2, \dots, X_n\}$, by:

$$f_{M_n}(x) = n(F_X(x))^{n-1} f_X(x)$$

We say that the distribution F is in the maximum domain of attraction of the distribution, H if the existence $c_n > 0, d_n \in R$, so that:

$$\lim_{n \rightarrow \infty} P \left(\frac{M_n - d_n}{c_n} \leq x \right) = \lim_{n \rightarrow \infty} F^n(c_n x + d_n)$$

$$= H(x) \Leftrightarrow \frac{M_n - d_n}{c_n} \xrightarrow{d} H$$

We know that Fisher-Tippett ([5]) and Gnedenko ([6]) has proved H will be one of three Extreme Value distributions: Gumbel, Fréchet, Weibull.

We given Generalized Extreme Value – GEV as:

$$G(x) = \exp \left\{ - \left(1 + \zeta \frac{x - \mu}{\sigma} \right)^{-\frac{1}{\zeta}} \right\};$$

In there: $1 + \zeta \frac{x - \mu}{\sigma} > 0; \mu \in R; \sigma > 0; \zeta \in R$.

We make the symbol Extreme Value distribution by GEV (μ, σ, ζ).

So that Extreme Value distributions has three parameters. Specify, we classify GEV by shape parameter ζ as:

✓ $\zeta = 0 \Rightarrow$ Gumbel distribution (EV I),

$$G(x) = \exp \left\{ - \exp \left(- \frac{x - \mu}{\sigma} \right) \right\}$$

✓ $\zeta = \frac{1}{\alpha} > 0 \Rightarrow$ Fréchet distribution (EV II),

$$G(x) = \exp \left\{ - \left(1 + \frac{1}{\alpha} \left[\frac{x - \mu}{\sigma} \right]^{-\alpha} \right) \right\}$$

✓ $\zeta = -\frac{1}{\alpha} < 0 \Rightarrow$ Weibull distribution (EV III),

$$G(x) = \exp \left\{ - \left[1 - \frac{1}{\alpha} \left[\frac{x - \mu}{\sigma} \right]^\alpha \right] \right\}$$

Characteristic of Extreme Value distribution with $\zeta \neq 0$.

The expectation: $E(X) = \mu - \frac{\sigma}{\zeta} + \frac{\sigma}{\zeta} g_1$

The variance of Gumbel distribution:

$$\text{Var}(X) = \begin{cases} \frac{\sigma^2}{\zeta^2} (g_2 - g_1^2); & \text{if } \zeta \neq 0 \text{ and } \zeta < \frac{1}{2}; \\ \infty; & \text{if } \zeta \geq \frac{1}{2} \end{cases}$$

there, $g_i = \Gamma(i)$, $i = 1, 2$ and $\Gamma(i) = \int_0^\infty t^{i-1} e^{-t} dt$

is gammafunction.

Characteristic of Gumbel distribution ($\zeta = 0$).

Cumulative distribution function (CDF):

$$F(x) = \exp \left\{ - \exp \left(- \left(\frac{x - \mu}{\sigma} \right) \right) \right\}.$$

The density function (PDF) of Gumbel distribution will be:

$$f(x) = \frac{1}{\sigma} \exp \left(- \left(\frac{x - \mu}{\sigma} \right) \right) \exp \left\{ - \exp \left(- \left(\frac{x - \mu}{\sigma} \right) \right) \right\}$$

The Expectation of Gumbel distribution:

$$EX = \mu + 0,577216 \sigma$$

there, $0,577216 \equiv$ Euler constant.

The Variance of Gumbel distribution:

$$\text{Var} X = \frac{(\pi\sigma)^2}{6}$$

In the next section of this review, application III will be presented the statistic methods Maximum Likelihood Estimation (MLE).

The method of moment is to find out the initial parameter $\mu^0; \sigma^0$ in Expectation of Gumbel Distribution, then MLE is used to estimate statistical parameters $\mu; \sigma$. The Newton-Raphson algorithm can be required to solve the set of equations.

2.2 The copula and distributions in the multivariate maximal domain of attraction.

Copula function is an efficient method for describing the dependence among multiple hydrologic variables. It offers a flexible way to construct a joint distribution independent from the marginal distributions. The main advantage of this approach is that construction of a joint distribution through a copula is independent of the marginal distributions of the individual variables.

Derived from the following definition of maximum suction region of multidimensional distribution:

Let $\{X, X_n, n \geq 1\}$ be i.i.d random vectors with common distribution F . Then X is in a multivariate maximal domain of attraction if there exist

$$a(t) = (a^{(1)}(t), \dots, a^{(m)}(t)) \in R_+^m$$

$$b(t) = (b^{(1)}(t), \dots, b^{(m)}(t)) \in R^m$$

such that

$$P^n \left(\frac{\mathbf{X} - \mathbf{b}(n)}{\mathbf{a}(n)} \leq \mathbf{x} \right) = F^n (\mathbf{a}(n)\mathbf{x} + \mathbf{b}(n))$$

$$= \left[P \left(\frac{X^{(i)}}{a^{(i)}(n)} \leq x^{(i)} \right); i = 1, \dots, m \right]^n \rightarrow G(\mathbf{x}) \quad (1)$$

Where G is a non-degenerate distribution called a max-stable or extreme value distribution. The marginal distributions $G_{\gamma^{(i)}}^{(i)}; i = 1, \dots, m;$ of G are one-dimensional extreme value distribution of the type

$$G_{\gamma^{(i)}}^{(i)}(x^{(i)}) = \exp \left\{ - \left(1 + \gamma^{(i)} x^{(i)} \right)^{-\frac{1}{\gamma^{(i)}}} \right\}; 1 + \gamma^{(i)} x^{(i)} > 0$$

and $G_{\gamma^{(i)}}^{(i)}(x^{(i)})$ concentrates on $\{u \in R : 1 + \gamma^{(i)} x^{(i)} > 0\}$.

Kluppelberg and Resnick ([4], 80 – 82) discuss aggregation of risk when (1) holds with $\gamma^{(i)} = 0$ for $i = 1, \dots, m;$ so that each marginal is in the domain of attraction of the Gumbel distribution. This is equivalent to supposing that, for $i = 1, \dots, m;$ there exists a self-neglecting function $f^{(i)}(t)$ with derivative converging to 0 such that

$$\frac{\bar{F}^{(i)}(t + xf^{(i)}(t))}{\bar{F}^{(i)}(t)} \rightarrow e^{-x}; x \in R$$

as t converges to the right endpoint $F^{(i)}; \bar{F}^{(i)}$ – joint survival function. Then we may take:

$$b^{(i)}(t) = \left(\frac{1}{1 - F^{(i)}(\cdot)} \right)^{-} (t); a^{(i)}(t) = f^{(i)}(b^{(i)}(t));$$

$$i = 1, \dots, m$$

There: $\varphi^-(t) = \inf \{y : \varphi(y) < t\}$.

To get attractive formulae, it is necessary to assume all marginal of F are the same so we proceed under the assumption:

$$F^{(i)}(\cdot) = F^{(1)}(\cdot), i = 1, \dots, m.$$

We now report the main result concerning risk aggregation: Suppose (1) holds where all marginal of $F(\mathbf{x})$ are equal and all marginal of $G(\mathbf{x})$ are Gumbel and (2), (3) hold.

Formulae for aggregation of risks may be readily obtained when F does not possess asymptotic independence.

3 APPLICATIONS

TABLE 1
WATER LEVEL

Step j	μ^j	σ^j	Δ_j	$< k = 10^{-4}$
0	379.9874	50.00598		
1	377.0478	57.715	51.516	$> 10^{-4}$
2	376.0177	64.818	15.47	$> 10^{-4}$
3	375.4537	68.71	0.7335	$> 10^{-4}$
4	375.3104	69.5544	0.00148	$> 10^{-4}$
5	375.3042	69.58668	4.1×10^{-4}	$> 10^{-4}$

3.1 Extreme Value distributions in the hydrological models.

Theory used in this application is shown in 2.1, and two specific assignments are solved as follows:

Assignment 1: Find maximum Gumbel distribution function for maximum rainfall in Tan Chau, An Giang based on data at hydrological stations in Tan Chau, An Giang, Vietnam.

Assignment 2: Find maximum Gumbel distribution function for maximum water level in

TABLE 2
RAINFALL

Step j	μ^j	σ^j	Δ_j	$< k = 10^{-4}$
0	73.4576	16.2605		
1	72.92	17.997	3.303	$> 10^{-4}$
2	72.725	18.777	0.647	$> 10^{-4}$
3	72.698	18.889	0.0134	$> 10^{-4}$
4	72.69766	18.891	3.8×10^{-6}	$< 10^{-4}$

Tien river based on data at hydrological station in Tan Chau, An Giang, Vietnam.

Data supplied by Southern Hydrographical Meteorological Station at the Tan Chau and Chau Doc districts, An Giang province from 1990 to present.

The problems are solved specifically as follows:

The expectation of Gumbel distribution:

$$EX = \mu + 0,577216 \sigma$$

there, $0,577216 \equiv$ Euler constant.

The variance of Gumbel distribution:

$$Var X = \frac{(\pi\sigma)^2}{6}$$

By moment method we will find the statistical estimation,

$$\begin{cases} \bar{X} = \mu + 0,577216 \sigma \\ S^2 = \frac{(\pi\sigma)^2}{6} \end{cases} \Rightarrow \begin{cases} \tilde{\mu} \approx \bar{X} - 0,4501 S \\ \tilde{\sigma} \approx 0,7797 S \end{cases}$$

there:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}; S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1};$$

The next section we will use maximum likelihood method for parameter estimation in distribution maximum to illustrate the use of actual data through Newton – Raphson algorithm.

First of all, to building the likelihood function, we have:

$$\begin{aligned} L(\mu, \sigma) &= f(x_1, x_2, \dots, x_n | \mu, \sigma) = \prod_{i=1}^n f(x_i | \mu, \sigma) \\ &= \prod_{i=1}^n \frac{1}{\sigma} \exp\left\{-\frac{x_i - \mu}{\sigma}\right\} \exp\left\{-\exp\left\{-\frac{x_i - \mu}{\sigma}\right\}\right\} \\ &= \left(\frac{1}{\sigma}\right)^n \exp\left\{-\sum_{i=1}^n \left[\frac{x_i - \mu}{\sigma} + \exp\left\{-\frac{x_i - \mu}{\sigma}\right\}\right]\right\}. \end{aligned}$$

$$\begin{aligned} \Rightarrow \Lambda &= \ln(L(\mu, \sigma)) = \\ &= -n \ln(\sigma) - \sum_{i=1}^n \left[\frac{x_i - \mu}{\sigma} + \exp\left\{-\frac{x_i - \mu}{\sigma}\right\} \right]. \end{aligned}$$

$$\begin{aligned} \frac{\partial \Lambda}{\partial \mu} &= -\sum_{i=1}^n \left[\frac{-1}{\sigma} + \frac{1}{\sigma} \exp\left\{-\frac{x_i - \mu}{\sigma}\right\} \right] = \\ &= \frac{n}{\sigma} - \frac{1}{\sigma} \sum_{i=1}^n \exp\left\{-\frac{x_i - \mu}{\sigma}\right\}. \end{aligned}$$

$$\begin{aligned} \frac{\partial \Lambda}{\partial \sigma} &= -\sum_{i=1}^n \left[\frac{-1}{\sigma} + \frac{1}{\sigma} \exp\left\{-\frac{x_i - \mu}{\sigma}\right\} \right] = \\ &= \frac{n}{\sigma} - \frac{1}{\sigma} \sum_{i=1}^n \exp\left\{-\frac{x_i - \mu}{\sigma}\right\}. \end{aligned}$$

$$\begin{aligned} \frac{\partial \Lambda}{\partial \sigma} &= -\frac{n}{\sigma} - \sum_{i=1}^n \left[-\frac{x_i - \mu}{\sigma^2} + \frac{x_i - \mu}{\sigma^2} \exp\left\{-\frac{x_i - \mu}{\sigma}\right\} \right] \\ &= -\frac{n}{\sigma} + \sum_{i=1}^n \frac{x_i - \mu}{\sigma^2} - \sum_{i=1}^n \frac{x_i - \mu}{\sigma^2} \exp\left\{-\frac{x_i - \mu}{\sigma}\right\}. \end{aligned}$$

We select $\hat{\mu}$ and $\hat{\sigma}$ which satisfactory the system of equations:

$$\begin{cases} \frac{\partial \Lambda}{\partial \mu} = 0 \\ \frac{\partial \Lambda}{\partial \sigma} = 0 \end{cases}$$

Using the Newton – Raphson algorithm:

$$\begin{aligned} \frac{\partial^2 \Lambda}{\partial \mu^2} &= -\frac{1}{\sigma} \sum_{i=1}^n \left(\frac{1}{\sigma} \right) \exp\left\{-\frac{x_i - \mu}{\sigma}\right\} = \\ &= -\frac{1}{\sigma^2} \sum_{i=1}^n \exp\left\{-\frac{x_i - \mu}{\sigma}\right\}. \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \Lambda}{\partial \mu \partial \sigma} &= -\frac{n}{\sigma^2} + \frac{1}{\sigma^2} \sum_{i=1}^n \exp\left\{-\frac{x_i - \mu}{\sigma}\right\} - \\ &= -\frac{1}{\sigma} \sum_{i=1}^n \frac{x_i - \mu}{\sigma^2} \exp\left\{-\frac{x_i - \mu}{\sigma}\right\} = \\ &= -\frac{n}{\sigma^2} + \frac{1}{\sigma^2} \sum_{i=1}^n \exp\left\{-\frac{x_i - \mu}{\sigma}\right\} - \\ &= -\frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu) \exp\left\{-\frac{x_i - \mu}{\sigma}\right\}. \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \Lambda}{\partial \sigma^2} &= \frac{n}{\sigma^2} - \frac{2}{\sigma^3} \sum_{i=1}^n (x_i - \mu) + \\ &= \frac{2}{\sigma^3} \sum_{i=1}^n (x_i - \mu) \exp\left\{-\frac{x_i - \mu}{\sigma}\right\} - \\ &= -\frac{1}{\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \exp\left\{-\frac{x_i - \mu}{\sigma}\right\}. \end{aligned}$$

Set:

$$f = \begin{bmatrix} \frac{\partial \Lambda}{\partial \mu} \\ \frac{\partial \Lambda}{\partial \sigma} \end{bmatrix}; K = \begin{bmatrix} \frac{\partial^2 \Lambda}{\partial \mu^2} & \frac{\partial^2 \Lambda}{\partial \mu \partial \sigma} \\ \frac{\partial^2 \Lambda}{\partial \mu \partial \sigma} & \frac{\partial^2 \Lambda}{\partial \sigma^2} \end{bmatrix};$$

$$\begin{bmatrix} \alpha^{(j+1)} \\ \sigma^{(j+1)} \end{bmatrix} = \begin{bmatrix} \alpha^{(j)} \\ \sigma^{(j)} \end{bmatrix} - K^{-1}(\alpha^{(j)}, \sigma^{(j)}) f(\alpha^{(j)}, \sigma^{(j)}); i = 0, 1, 2, \dots$$

This iteration is repeated until the following inequality is satisfied:

$$\left\| \begin{bmatrix} \mu^{(j+1)} \\ \sigma^{(j+1)} \end{bmatrix} - \begin{bmatrix} \mu^{(j)} \\ \sigma^{(j)} \end{bmatrix} \right\|^T \left\| \begin{bmatrix} \mu^{(j+1)} \\ \sigma^{(j+1)} \end{bmatrix} - \begin{bmatrix} \mu^{(j)} \\ \sigma^{(j)} \end{bmatrix} \right\| < k.$$

$$\Delta_j = \left(\mu^{(j+1)} - \mu^{(j)} \right)^2 + \left(\sigma^{(j+1)} - \sigma^{(j)} \right)^2 < k.$$

The problems for the maximum distribution of water lever (or rainfall) in Tan Chau station, if select $k = 10^{-4}$.

3.2 WATER LEVEL PROBLEM

Based on data: $\mu^0 = 379.9874$; $\sigma^0 = 50.00598$ calculated by moment method in the previous method of maximum likelihood and Newton – Raphson algorithm, we get the results in repeated step as table below.

The extreme distribution function in this problem will have:

$$F(x) \approx \exp \left\{ -\exp \left[\frac{-(x - 375.3042)}{69.59} \right] \right\}$$

3.3 RAINFALL PROBLEM

Based on data: $\mu^0 = 73.4576$; $\sigma^0 = 16,2605$ calculated by moment method in the previous method of maximum likelihood and Newton – Raphson algorithm, we get the results in repeated step as table below

The extreme distribution function in this problem will have:

$$F(x) \approx \exp \left\{ -\exp \left[\frac{-(x - 72.69766)}{18.891} \right] \right\}$$

4 CONCLUSION

Rainfall in Tan Chau, An Giang and Tien River through Tan Chau station, An Giang, are two random variables that are generally non-independent but dependent on each other. Therefore, when considering the extreme problem of water volume in general, it is necessary to consider Copula among them

The problems outlined in this paper may be more general in that they include multivariate processes (floating and underground flows, tidal impacts, two seasons of the South, rainfall and In fact, we have (see [2],[7]) and will study the above issues to clarify the problems of hydrological hydrology in the South.

REFERENCES

- [1] Enrique Castillo, Ali S. Hadi, N. Balakrishnan, Jose Maria Sarabia .“Extreme Value Models with Applications in Engineering and Science”,193-207. Wiley-Interscience, 2005.
- [2] Duong Ton Dam, Huynh To Uyen, Duong Ton Thai Duong.”Extreme Value Distributions Class and Applications in Hydrological Analysis and Risk Management”International Conference in Mathematics and Applications Mahidol University.ICMA-MU 2016 Book on the Conference Proceedings,39-48 (CEM), 2016
- [3] Umberto Cherubini, Fabio Gobbi, Sabrina Mulinacci, Silvia Romagnoli. “Dynamic Copula Methods in Finance”. John Wiley & Sons. Ltd, 2012.
- [4] Claudia Kluppelberg, Cidney I. Resnick. “The Pareto Copula, Aggregation of Risks, and The Emperor’s Sosks” J.Appl. Prob. 45, 76-84, 2008.
- [5] Fisher, R. and Tippett, L.H.C. “Limiting forms of the frequency distribution of largest or smallest member of a

sample”. Proceedings of the Cambridge Philosophical Society, 24: 180-190 ,1928.

- [6] Gnedenko, B. V. “Sur la distribution limmite du terme d’une serie aleatoire” Annals of Mathematics, 44:423-453, 1943..
- [7] Duong Ton Dam, Vo Van Tai, Pham Minh Truc, Dang Kien Cuong. “Forecasting crest of sanility at three main stations of CaMau province by fuzzy time series model”. Can Tho University Journal of Science, 47: 68-78, 2016.

Duong Ton Dam was born in Vietnam in 1952. He received the Ph.D degree in mathematics from Russia. He was Vice Rector of University of Information Technology, Vietnam National University, Ho Chi Minh City 2007 – 2012.

He is currently General Secretary of the Mathematical Society Ho Chi Minh City.

Dang Kien Cuong was born in Nam Dinh, Vietnam in 1972. He received the B.S.degree in mechanical engineering from Nong Lam University and Information Technology from Hanoi University of Science and Technology, M.S degree in computer science from AIT. He is currently pursuing the Ph.D. degree in computer science University of Information Technology, Vietnam National University, Ho Chi Minh City. He is currently the lecture at NongLam University, Ho Chi Minh City.

Sử dụng lý thuyết về quá trình ngẫu nhiên để giải bài toán phân tích dữ liệu thủy văn

Dương Tôn Đàm, Đặng Kiên Cường

Tóm tắt - Trong bài báo xét đến bài toán phân tích dữ liệu thủy văn (lượng mưa, dòng chảy, mực nước,...) để đánh giá, dự báo về những đợt biến như: Đỉnh lũ, cực hạn, cực mặn,... nhằm hạn chế những tác hại của biến đổi khí hậu lên kinh tế và đời sống. Công cụ chính để giải quyết bài toán đặt ra là – phân phối cực trị và miền hút max của chúng cùng với việc xét Gumbel copula của các biến ngẫu nhiên liên quan đến dữ liệu thủy văn. Các kết quả nêu ra trong bài báo được xem xét và minh chứng qua các dữ liệu của tỉnh An Giang tại các trạm thủy văn Tân Châu và Châu Đốc từ năm 1990 đến nay.

Từ khóa - copula; phân phối cực trị (Extreme Value distributions); miền hút max (Maximal domain of attraction); phương pháp hợp lý cực đại (Method of maximum likelihood); thuật giải Newton – Raphson (Newton – Raphson algorithm)